

Tools gen-1. Common Assessment Terminology (p. 1 of 2)

Common Assessment Terminology

Alignment — The process of linking content and performance standards to assessment, instruction, and learning in classrooms.

Confidence interval — A range of values constructed from sample data so that a population parameter occurs within that range at a pre-selected probability. A confidence interval (or confidence band) is an area of statistical similarity. Sometimes considered as a “buffer”, the confidence interval reflects an area (or range) within which one group (e.g., the students in a school) is considered NOT to be statistically significantly different from another group (e.g., the students in the state).

Construct validity — A test has construct validity when it actually measures the knowledge domain or behavior it claims to measure. For instance, if you give a social studies test and a student does poorly because the reading level was too difficult. That test does not have construct validity, because it was really measuring reading ability.

Content validity — A test has content validity if it adequately samples the knowledge or behavior that has been the goal of instruction.

Criterion-referenced Testing — Taking student scores on an instrument and comparing them to a standard. Criterion-referenced assessment tells us how well students are performing on specific goals or standards rather than how their performance compares to a norm group of students nationally or locally.

Cross-tabulate — Compare subgroups of students on two or more variables. Use multiple sources of data about a group of individuals to search for patterns or variations.

Cross-sectional — Studying different groups of students at a particular age or stage of development. Example: Measuring fourth grade students each year on ITBS and comparing the results.

Disaggregate — Partition the data by relevant subgroups (e.g., race, ethnicity, or gender).

Effect Size — A way of quantifying a difference between repeated measures. From a research perspective, there are difficulties associated with pretest to posttest contrasts. However, from an “in the busy classroom” perspective, such contrasts usually help teachers arrive at sensible conclusions about the effectiveness of their instruction (Popham, 1999). One method for calculating effect size is shown as:

$$\text{Effect Size} = \frac{\text{Mean of Posttest} - \text{Mean of Pretest}}{\text{Standard Deviation of Pretest}}$$

Longitudinal — Studying the same group of students over a period of time.

Mean — The average of the scores. Example: Thirty students in a class take a test. To get the mean score, add up the raw scores of all 30 students and divide by 30.

Median — The point in a distribution above which and below which 50% of the scores lie. Example: 11 students take a test. Their scores are 100, 98, 95, 94, 92, 88, 86, 86, 85, 83, 77. The median is 88 because it is the middle score.

Tools gen-1. Common Assessment Terminology (p. 2 of 2)

Mode — The score obtained by the largest number of individuals taking a test. Example: 11 students take a test. Their scores are 100, 98, 95, 94, 92, 88, 86, 86, 85, 83, 77. The most frequent score is 86. The mode is 86. In the case of two most frequently occurring scores, such as 100, 98, 95, 94, 92, 92, 88, 86, 86, 83, 77, there are two modes (bimodal): 92 and 86. A multimodal test has a distribution with more than 2 modes.

Normal distribution curve — A theoretical curve noted for its bell-shaped form. In a normal distribution about 68% of the values are within one standard deviation of the mean and about 95% of the values are within 2 standard deviations of the mean.

Norm-referenced testing — Taking student scores on an instrument and comparing those scores to other students or a normed group of peers. ACT, SAT, and GRE are examples of norm-referenced tests.

Percentile score — A score that tells the percentage of individuals taking a test who are at or below a particular score; a percentile rank of 85, for example, means that the student did as well or better than 85% of those taking the test.

Qualitative data — Observations that consist of words, labels, or numerical codes.

Quantitative data — Observations consisting of numbers that indicate differences in amounts.

Range — Indicates the measure of variability from the highest score to the lowest score.

Reliability — A test is said to be reliable if a student's scores would be the same in repeated testing.

SES — Socioeconomic status characteristic(s). Whether or not a student receives free or reduced priced lunches is often used as a measure of SES.

Standard deviation — A measure of variability that indicates the how scores differ from the mean or average score.

Validity — A test is said to have validity if it measures what it is supposed to measure and inferences and actions made on the basis of test scores are appropriate and accurate. For example, if a student performs well on a reading test, how confident are we that that student is a good reader? A valid standards-based assessment is aligned with the standards intended to be measured, provides an accurate and reliable estimate of students' performance relative to the standard, and is fair. Further, an assessment cannot be valid if it is not reliable.

Tools gen-2. Acronyms and Abbreviations (one page)

Acronyms and Abbreviations

For convenience, this list is included at the beginning of Part 1.

AEA	Area Education Agency
AIG	Annual Improvement Goals
AMO	Annual Measurable Objective
AYP	Adequate Yearly Progress
CSIP	Comprehensive School Improvement Plan
DCDP	District Career Development Plans
DDL	Data Driven Leadership
DE	Department of Education
ELL	English Language Learner
FR	Free/Reduced Lunch
IASB	Iowa Association of School Boards
ITBS	Iowa Test of Basic Skills
ITCDP	Individual Teacher Career Development Plan
ITED	Iowa Test of Educational Development
LEA	Local Education Agency
LEP	Limited English Proficient
NAEP	National Assessment of Educational Progress
NSCD	National Staff Development Council
NCLB	No Child Left Behind
NPR	National Percentile Rank
PD	Professional Development
SE	Special Education
SES	Socio-Economic Status
SINA	School in Need of Assistance

Tools gen-3. Four Operating Principles (one page)

Four “Operating Principles”



Focus on Curriculum, Instruction, and Assessment

A clear focus on instruction is essential. Deliberate alignment of instruction, curriculum, and assessment increases the likelihood that professional development efforts will be effective. If the goal is increased student achievement, use the most powerful tools over which the school has control.

Participative Decision Making

Collective action requires a democratic process. Teachers are engaged in decision making and planning for professional development that is aligned with identified student needs. Communication and governance processes are in place to increase the likelihood that decisions made about staff development are binding. When professional development decisions affect a group (rather than an individual), group members must participate in those decisions.

Leadership

Strong leaders are essential for successful professional development efforts. Leaders facilitate the engagement of all faculty members responsible for instruction, address time and resource issues and balance both the pressure and support required to sustain professional development efforts as a priority. For leadership to be pervasive and intense enough to make things happen at the district, building, and classroom levels, it must be distributed through the organization – involving the school board, central office administration, building-level administration, and teachers. Collective professional development aimed at student learning goals requires focused leadership.

Simultaneity

Schools and districts often have to attend to multiple concerns simultaneously. Professional development efforts balance the resources directed toward and the efforts invested in content, context and processes. To accomplish student achievement gains, focusing on new content is the priority but simultaneously issues of context and process may also need to be addressed. Select a priority in which to invest professional development time and resources and then seek ways to integrate other concerns without losing focus on the major initiative. If multiple initiatives receive equal effort, the probability of succeeding with any of them is reduced.

